國立交通大學

資訊科學與工程研究所

碩 士 論 文

基於賈伯小波之自動人臉年齡偵測

Automatic Age Estimation of Face Image based on Gabor Wavelets

研 究 生：賴健豪

指導教授：許騰尹 教授

林進燈 教授

中華民國 九十九 年 八月

基 於 賈 伯 小 波 之 自 動 人 臉 年 齡 偵 測

Automatic Age Estimation of Face Image based on Gabor Wavelets

研 究 生：賴健豪　　　　　　　　　　Student：Jian-Hao Lai

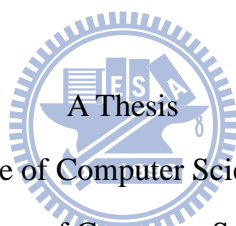指導教授：許騰尹 教授　　　　　　　Advisor：Prof. Terng-Yi Hsu

　　　　　林進燈 教授　　　　　　　　　　Prof. Chin-Teng Lin

國 立 交 通 大 學

資 訊 科 學 與 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

August 2010

Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 九 年 八 月

基於賈伯小波之自動人臉年齡偵測

學生：賴健豪　　　　　　　　指導教授： 許騰尹 教授

林進燈 教授

國立交通大學資訊科學與工程研究所

摘要

近年來，年齡偵測在人臉辨識技術中，漸漸成為一項受重視的領域，加上科技的發達，年齡偵測被認為在多媒體通訊、人機介面、居家照護及安全監控等應用上有相當大的發展潛力。在本篇論文中，我們提出一個創新與可靠的自動化人臉年齡偵測架構。本篇論文提出藉由結合賈伯小波與正交區域保留投影(Orthogonal Locality Preserving Projections)，產生適用於年齡辨識的新人臉特徵。接下來，在考量到同一個體各年齡層的影像資料蒐集的難易度以及現行公開之年齡資料庫的資料稀少，為了獲得更佳的一般性，我們採用基於支持向量機(Support Vector Machines)的分類器結合新人臉特徵。此系統可以全自動即時提取人臉特徵，相較於大多數文獻所提出的半自動人臉特徵提取，更具有實際應用的潛力。

本篇論文的目標是建構一個全自動且即時分析的年齡偵測系統，由此系統所得到的實驗結果將可以提供年齡偵測領域上更深入的了解，並且有助於實際應用的開發。

Automatic Age Estimation of Face Image based on Gabor Wavelest

Student: Jian-Hao Lai   Advisor: Prof. Terng-Yi Hsu

Prof.Chin-Teng Lin

Institute of Computer Science and Engineering

National Chiao Tung University

ABSTRACT

In recent years, age estimation has become an important research topic in face recognition technology. Furthermore, age estimation is considered as a potential research which has lots of real-world potential applications such as multimedia communication, human computer interaction, and security. In this thesis, we present a novel and reliable framework for automatic age estimation. It exploits the whole new face feature based on the combination of Gabor wavelets and Orthogonal Locality Preserving Projections. In order to obtain more proper generalization ability with respect to sparse training samples, we use a support vector machine based classifier. Since this system can extract face aging features automatically in real-time, it has more potential in applications than other semi-automatic systems.

The objective of this paper is to build a full automatic and real time age estimation system. The results obtained from this novel approach would provide better insight to operators in the field of age estimation to develop the real-world applications

致　　謝

　　在這研究期間當中，首先要感謝我的指導教授 許騰尹教授及共同指導教授 林進燈教授，在老師豐富的學識，使我學習到許多寶貴的知識，在面對事情中應有的處理態度，方法。在研究的過程之中，老師對於學問深入的看法，清晰的思緒都是我該學習的目標。
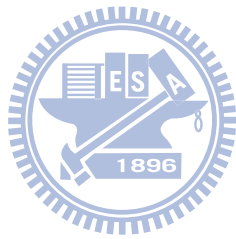
　　由衷感謝口試委員陶金旺教授，蒲鶴章博士給予許多寶貴的建議與指正，使得這篇論文更加完整。特別感謝東霖學長與肇廷學長，在這期間給予很多建議，在我研究上遇到困難時有很大的幫助。感謝全體同學佳芳，嶸健，聖傑與哲男在研究的過程之中不斷的互相討論，互相支持鼓勵一起分享不同的經驗，使得我更加成長。也感謝實驗室的學弟們，在完成論文時給予許多幫助。感謝一路陪伴在我身邊的所有朋友，使得我的研究生涯變得多采多姿。

特別要感謝我的父親，母親，在這段日子中不斷地給予我支持與鼓勵，提供我良好的生活環境使我能更專心在研究之上並且順利完成學業。最後誠摯地以本論文研究成果獻給我的師長，父母，家人，及所有的朋友們。

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Human face image contains massive information of personal characteristics including identity, emotion expression, gender, and age ...etc. Generally speaking, human face image can be considered as a complex signal which is composed by many facial attributes such as skin color, the geometric features of facial features, and so on. This attributes play a crucial role in real applications of facial image analysis; For example, multimedia communication, human computer interaction (HCI), and security. In such applications, various attributes that were estimated from a captured face image can infer the further system reactions. Age in particular is more significant among these attributes. For example, users may need an age specific human computer interaction system which can estimate their age for secure system access control or intelligence service. For secure system access control, the system ensures young kids have no access to enter the websites with adult materials. For intelligence service, a vending machine that secured by the system can refuse to sell cigarettes and alcohol drink to the underage people; ad-agency can receive the feedback of advertising effectiveness to change the the advertisements appropriate for the age range of potential customers in real-time. Obviously, automatic human age estimation by facial image analysis has lots of potential real-world applications.

Although automatic age estimation of face image is an important technique involved in many real-world applications, estimating human age from face images are

still a challenging problem. Because aging progress represents differently not only between races but also within races, which means the progress is almost personal; moreover, this progress is also determined by external factors, such as health, living style, living location, and weather conditions. Therefore, "How to find a robust representation featuring" this progress still remains an open problem. In this thesis, a full automatic age estimation system using Gabor wavelets to represent aging progress is proposed.

1.2 Related Works
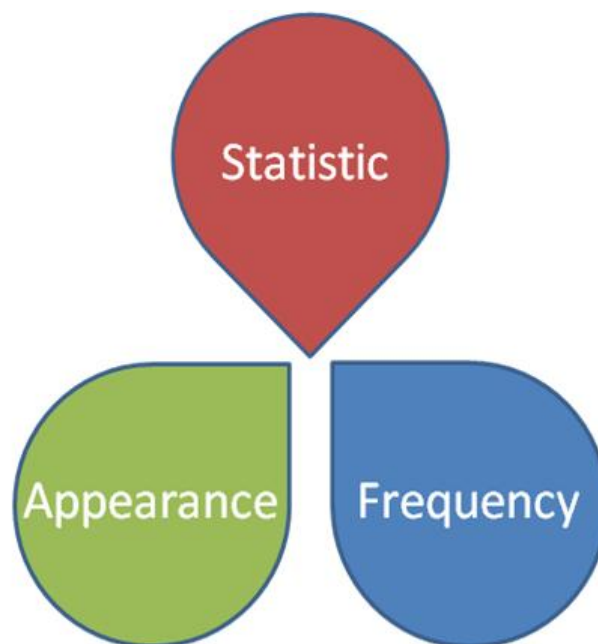
Fig.1-1: Three categories of human face age estimation

There are three categories of feature extraction for human face age estimation in the literature as shown in Fig. 1-1. The first category is *Statistical-Based* approaches. Xin Geng *et al.* [1][2] proposed the *AGing pattErn Subspace* (AGES) method (as shown in Fig. 1-2) for automatic age estimation. The idea of AGES is to model the

aging pattern, which is defined as a sequence of personal aging face images, by learning a representative subspace from EM-like (Expectation-Maximization) iterative learning *Principle Component Analysis* (PCA). In another major studies [3][4], Guodong Guo *et al.* compared three typical dimensionality reduction and manifold embedding methods such as PCA, *Locally Linear Embedding* (LLE) and *Orthogonal Locality Preserving Projections* (OLPP). According to the data distribution in OLPP subspace, they proposed the *Locally Adjusted Robust Regression* (LARR) method for learning and prediction of human ages. The concept of LARR is shown in Fig. 1-3. The LARR applies *Support Vector Regression* (SVR) to get a coarse prediction and find a local adjustment within a limited range of ages centered at the predicted result by *Support Vector Machine* (SVM).
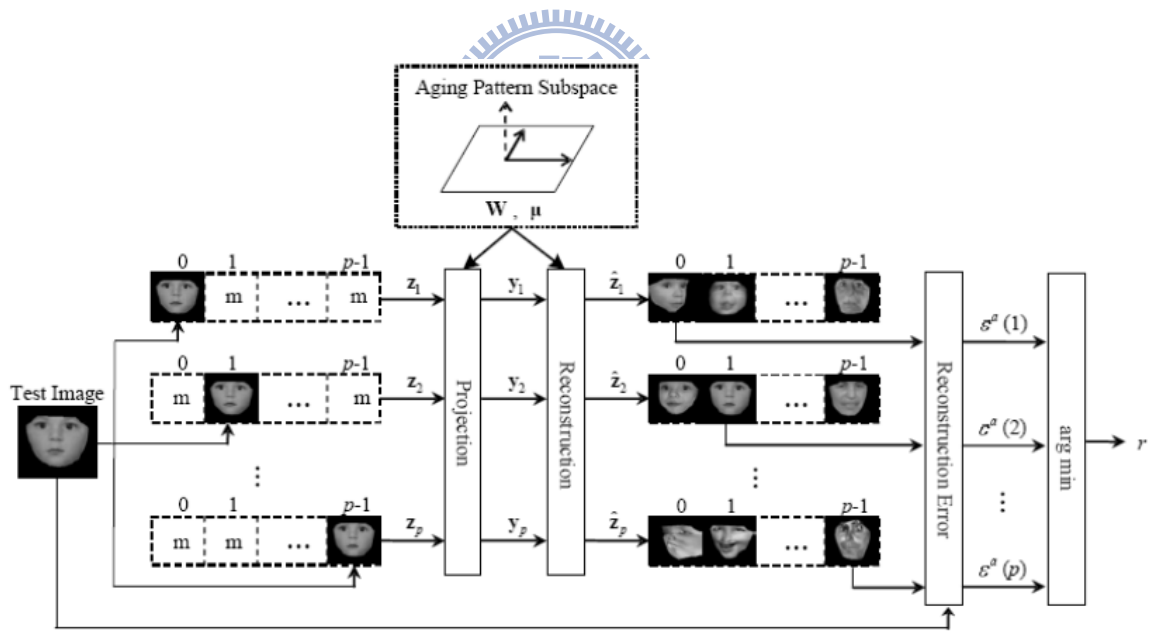
Fig. 1-2: Flowchart of AGES for age estimation in [1]
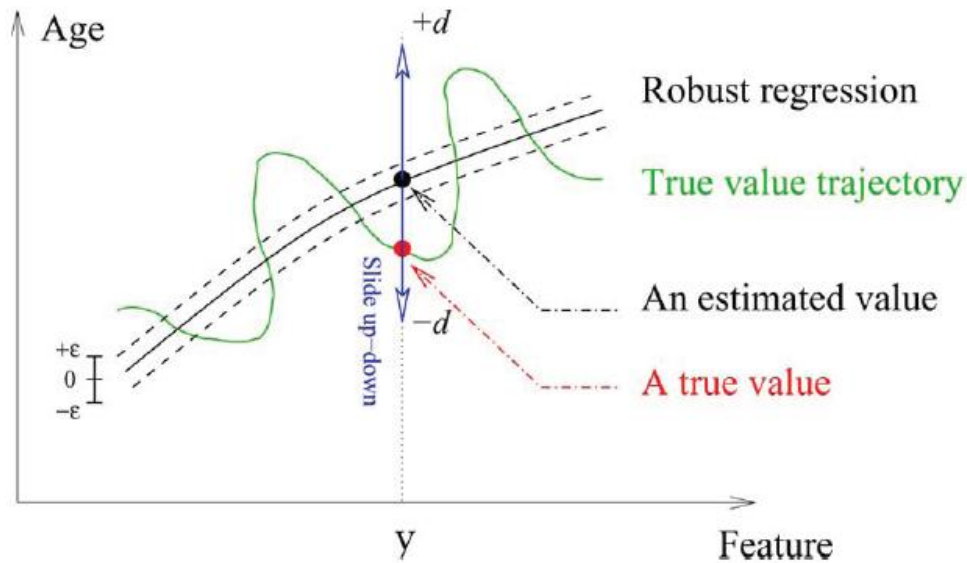
Fig. 1-3: Illustration of the concept of LARR in [3]

The second category is *Appearance-Based* approaches. Using appearance information is the most intuitial method in all facial image analysis works. Young H. Kwon *et al*. [5] used some visual aging features to construct an anthropometric model. The primary features are the eyes, nose, mouth, chin, etc. The ratios of those features are computed to distinguish different age ranges. In secondary feature analysis, a wrinkle geography map is used to guide the detection and measurement of wrinkles. Jun-Da Txia *et al.* [6] proposed an age estimation method by *Active Appearance Model* (AAM) to extract the regions of age features. Each face requires 28 feature points (see Fig. 1-4a) and is divided into 10 wrinkle feature regions (see Fig. 1-4b). The flowchart of this system is shown in Fig. 1-5. Shuicheng Yan *et al*. [7] present a patch-based appearance model named *Patch-Kernel*. This method is designed for characterizing the *Kullback-Leibler* divergence between the models which are derived from global *Gaussian Mixture Model* (GMM) using *Maximum a Posteriori* (MAP) for any two images.Its discriminating power is further enhanced by a weak learning process, called inter-modality similarity synchronization. Kernel regression is employed for estimate human age finally.
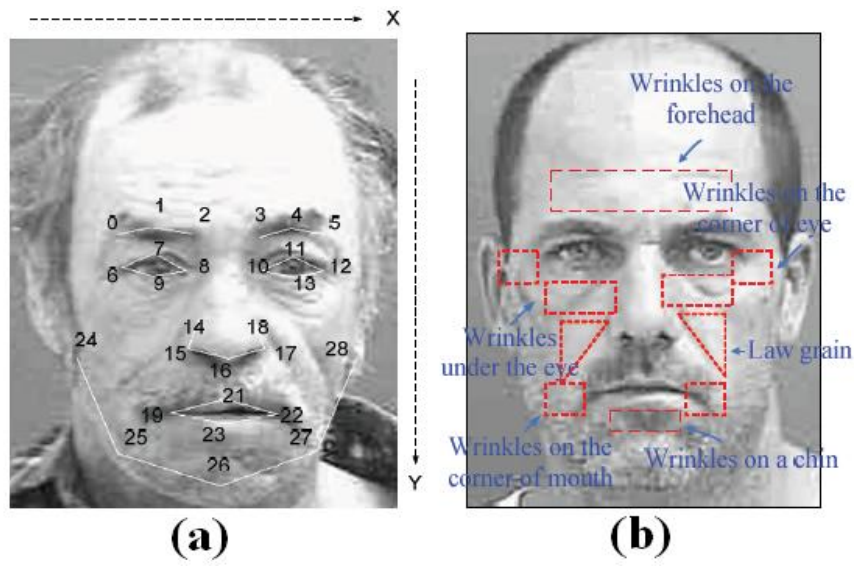
Fig. 1-4: In [6], (a) shows the locations of 28 feature points (b) shows the 10 wrinkle

regions
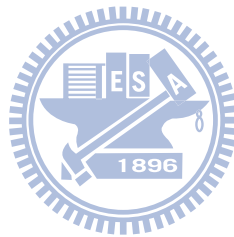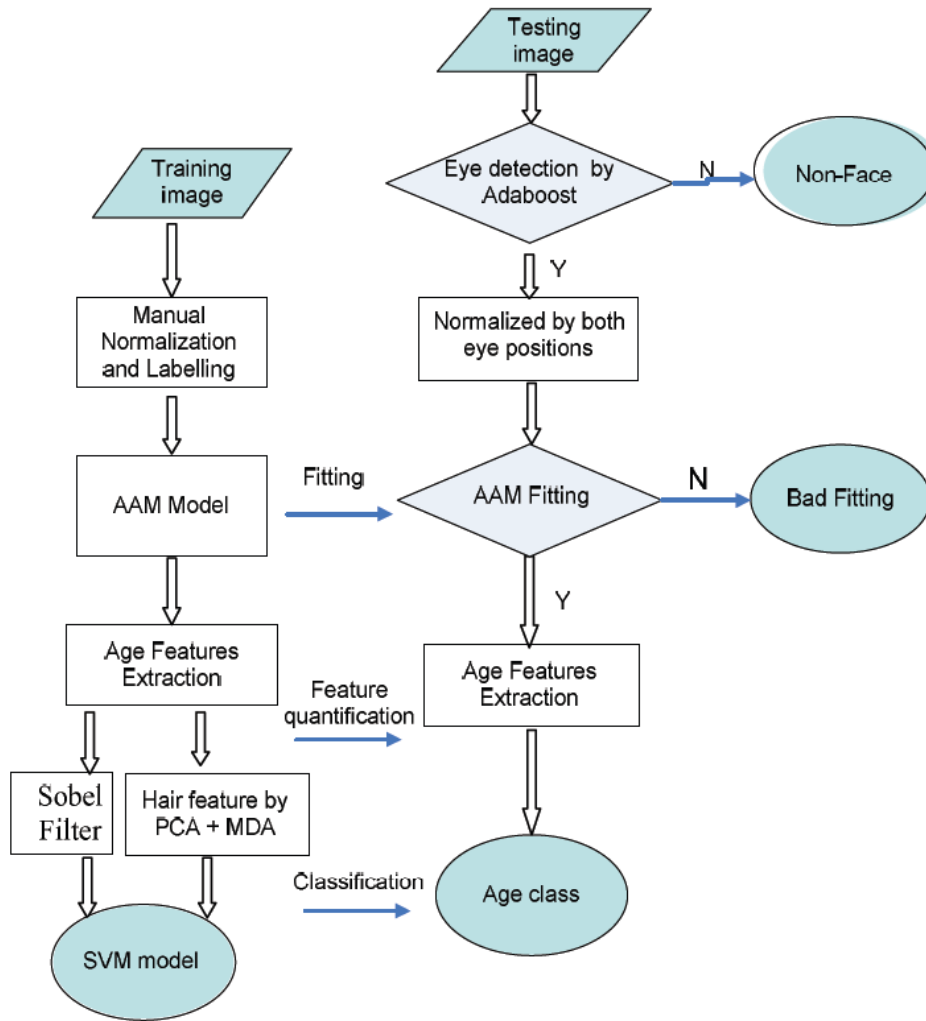
Fig. 1-5: Flowchart of age estimation system in [7]

The third category is *Frequency-Based* approaches. In image processing and pattern recognition, frequency domain analysis is the most popular method to extract the features. Guodong Guo *et al.* [8] investigate the *Biologically inspired Features* (BIF) for human age estimation from faces. Unlike the previous works done in [3][4], Guo simulated human visual process based on bio-inspired models [9] by applying Gabor filters. A Gabor filter is a linear filter used in image processing for edge detection. Frequency and orientation representations of Gabor filter are similar to those of human visual system, and it has been found to be particularly appropriate for texture representation and discrimination. Furthermore, previous bio-inspired models

are changed by proposing a novel "STD" operation.

1.3 System Overview

The system we proposed have four main modules: 1) face detection, 2) Gabor wavelets analysis, 3) OLPPs Reduction, and 4) SVM Classification as shown in the diagram of Fig. 1-6. The input image comes from camera frame or image file. First, face is captured from image by face detector which is achieved by using AdaBoost approach presented in [10] and the image is resized to 64 by 64 pixels. After face detection, using 40 Gabor wavelet kernels to extract features and reducing features by OLPPs. Last, estimating age form features by SVM classification. The detail of system implementation is further described in chapter 2 and chapter 3.
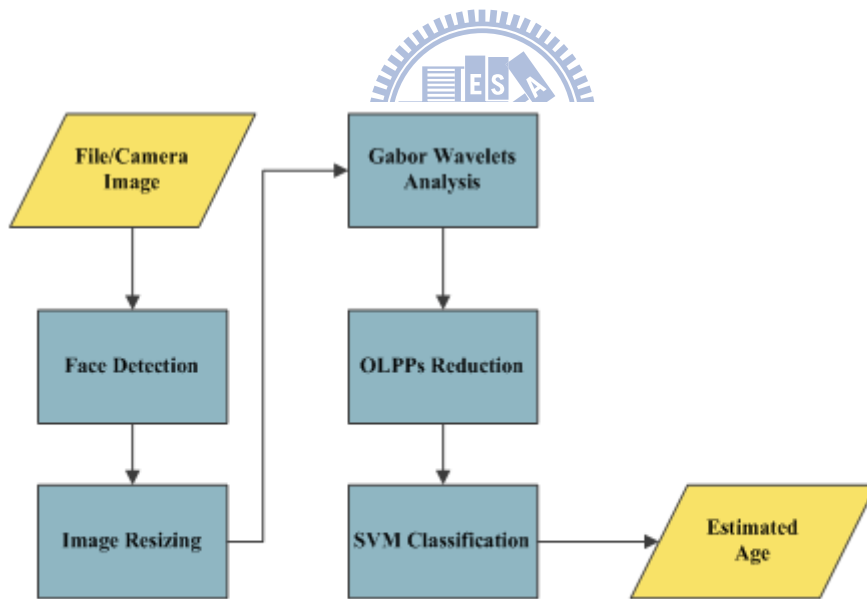
Fig. 1-6: System overview

1.4 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 describes

subsystem of face detection by AdaBoost. Chapter 3 shows face age estimation algorithm including texture analysis using Gabor wavelets, data reduction based on orthogonal locality preserving projections, and classification. Chapter 4 shows experimental results and comparisons. Finally, the conclusions of this system and future work will be presented in chapter 5.
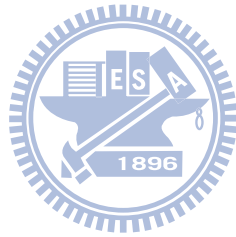
Chapter 2
Face Detection

2.1 Face Detection by AdaBoost

In this section, the face detection subsystem is introduced. The detection technique employed in this subsystem based on the AdaBoost algorithm that Viola and Jones proposed in 2004 [11]. The Viola-Jones classifier employs AdaBoost at each node in the cascade to learn a high detection rate at the cost of low rejection rate multitree classifier at each node of the cascade. The algorithm incorporate several innovative feature: 1) It uses Haar-like input features: a threshold applied to sums and differences of rectangular image regions. 2) Its integral image technique enables rapid computation of the value of rectangular regions or such regions rotated 45 degree. This data structure is used to accelerate computation of the Haar-like input features. 3) It uses statistical boosting to create binary (face-not face) classification nodes characterized by high detection and weak rejection. 4) It organizes the weak classifier nodes of a rejection cascade. In other words: the first group of classifiers is selected that best detects image regions containing an object while allowing many mistaken detections; the next classifier group is the second-best at detection with weak rejection; and so forth. In test mode, an object is detected only if fit it makes it through the entire cascade.

Fig. 2-1 shows the processes of face detection in implementation. Firstly, the searching windows with various sizes, which would be used to find different face candidates from the input images in multiscales. The face candidates in different scales truly reflect various distances of clients from input image. We totally define 12

searching windows in various sizes from the smallest block size of 24 by 24 to the biggest one by a multiplier of 1.25. All face candidates are normalized to same size and luminance. Then, Each normalized candidates will be classified as face or not by the face detector which was trained beforehand. The face regions are located precisely by a region-based clustering method in the end of the subsystem.
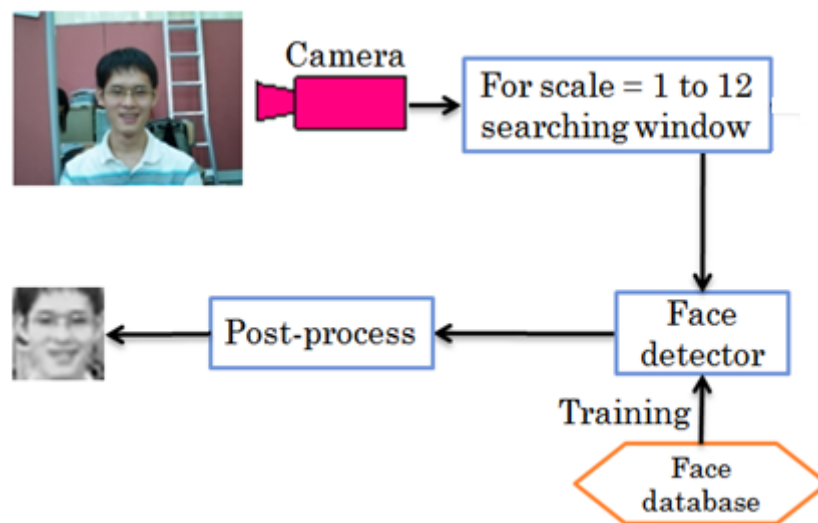
Fig. 2-1: Flowchart of face detection

2.2 Feature Selection

The intensity based features employed in this work were based on Haar features. Haar-like features are reminiscent of Haar basis functions which have been used by Papageorgiou [12]. In fact, it is feasible to use composition of different brightness rectangles to present the light and dark region in the image. The target object can be detected by contrasting unknown objects with the rectangle features if we can know the entire rectangle features which present the target object. Fig. 2-2 shows the difference in intensity between the region of the eyes and a region across the upper cheeks for charactering faces.

Fig. 2-2: The two of multiple rectangle features appear the face
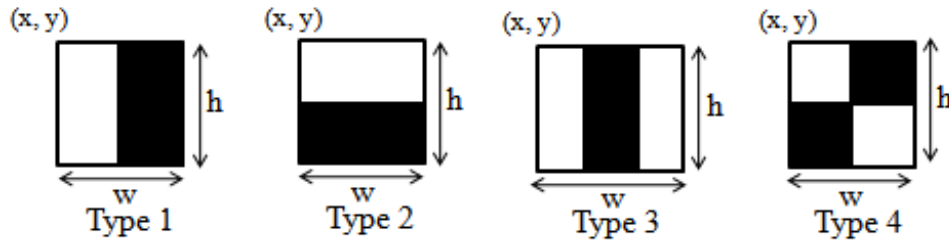
Fig. 2-3: The four kinds of rectangle features

We selected four types of rectangle features as illustrated in Fig. 2-3, including the vertical edge, horizontal edge, vertical line, and diagonal edge. The features are defined as Eq. (2.1).

$$valve_{subtracted} = f\left(x, y, w, h, Type\right) \tag{2.1}$$

where (*x, y*) is the origin of the relative coordinate of rectangle features in the searching window. The searching window is used to find the block which has a face inside in the image. The significance of *w* and *h* denotes the relative weight and height of rectangle features, respectively. *Type* represents the kind of rectangle feature. $valve_{subtracted}$ is the sum of the pixels in white rectangle subtracted from those in black rectangle.

A weak classification can be achieved by a single rectangle feature which best separates face and non-face images. That is, for each rectangle feature, the weak classification determiners the optimal threshold classification function, such that the

minimum number of examples are misclassified.

Fig. 2-4: The positive database

Fig. 2-5: The negative database
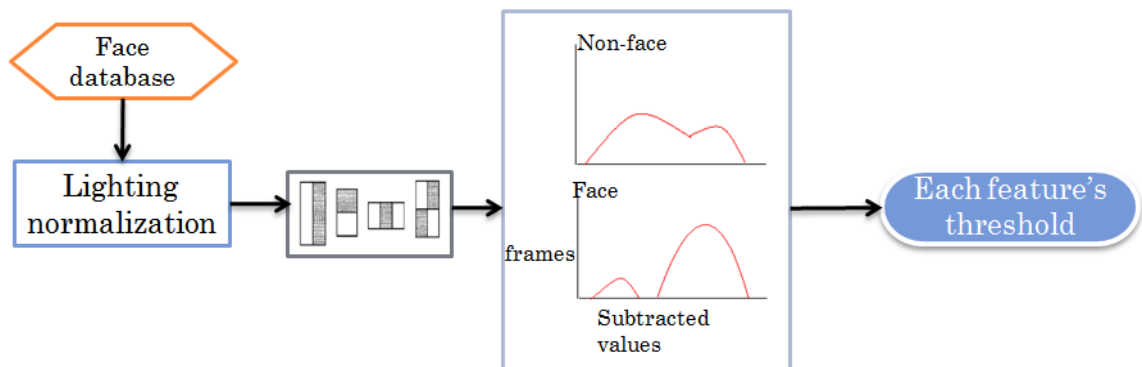
Fig. 2-6: The flow chart of selecting threshold for rectangle features

The selected threshold for each rectangle feature is acquired through the training

process by our database which contains 4,000 face images and 59,000 non-face

images. Fig. 2-4 and Fig. 2-5 present some face and nonface examples in our database.

The threshold selection process is shown in Fig. 2-6. In order to get the thresholds,

collecting the distribution information of subtracted values by this rectangle feature foe face database is needed. Find out a threshold which discriminates two classes to make detection rate higher than others. A single rectangle feature which best separates the face and nonface samples can be considered as a weak classifier $h(x, f, p, \theta)$ as shown in the following equation

$$h(x, f, p, \theta) = \begin{cases} 1, & if \ pf(x) < p\theta \\ 0, & otherwise \end{cases} \qquad (2.2)$$

This equation used to determine if the x-block image is a face or a nonface depends on the feature $f(x, y, w, h, type)$, a threshold (θ) and a polarity (p) indicating the signs of inequality. For each weak classifier, an optimal threshold is chosen to minimize the possibilities of misclassifications.. In this procedure, we could collect the distributions of $f(x, y, w, h, type)$ for each image in the database, and then a threshold with higher distinguishability in clustering would be chosen. Although each rectangle feature can be obtained easily, computing the complete set of all features is extremely costing. Take the smallest searching window of 24 by 24 block size, for example, the entire number of rectangle features will be 160,000.

2.3 Training of Face Detector

Viola and Jones [11] present a variant of AdaBoost is used both to select the rectangle features and to train the classifier. The origin form of AdaBoost learning algorithm is used to boost the classification performance of a single learning algorithm by combining a collection of weak classification functions to form a strong classifier.

Since the stronger classifier is rather time consuming, the structure of cascaded

classifiers by Viola and Jones [11] will be preferred to improving the detection performance and reducing the computational time. As a result, our cascaded Adaboost classification based on the stronger classifier will classify each extracted face image step by step as shown in Fig. 2-7.
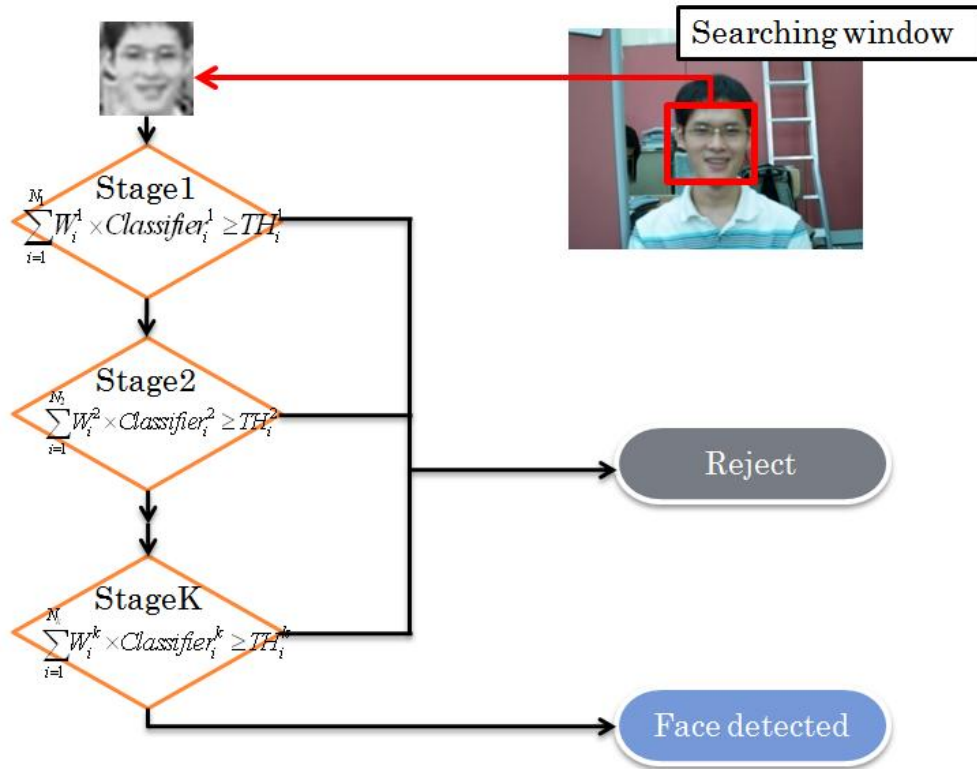
Fig. 2-7: The overall classifier

In stage1, an object extracted by searching window is classified as face so that it is allowed entering to stage2, otherwise the object is rejected. As same as in stage3 the object has to pass by stage2. In brief, a labeled face is passed through a series of stages; a rejected object is rejected by particular stage even if it enters the last stage.

Fig. 2-8 shows the flowchart of training of cascaded classification. f is the maximum acceptable false positive rate each stage, d is minimum acceptable detection rate each edge, $F_{t\arg et}$ is overall false positive rate, P is the set of face

samples, and N is the set of non-face samples. ^{i} means the stage of cascaded classification and ^{n_i} means the number of weak classification in the stage. The overall false positive rate must be smaller than F_{target} and each stage have to satisfy the equality: $F_i \leq f \times F_{i-1}$.
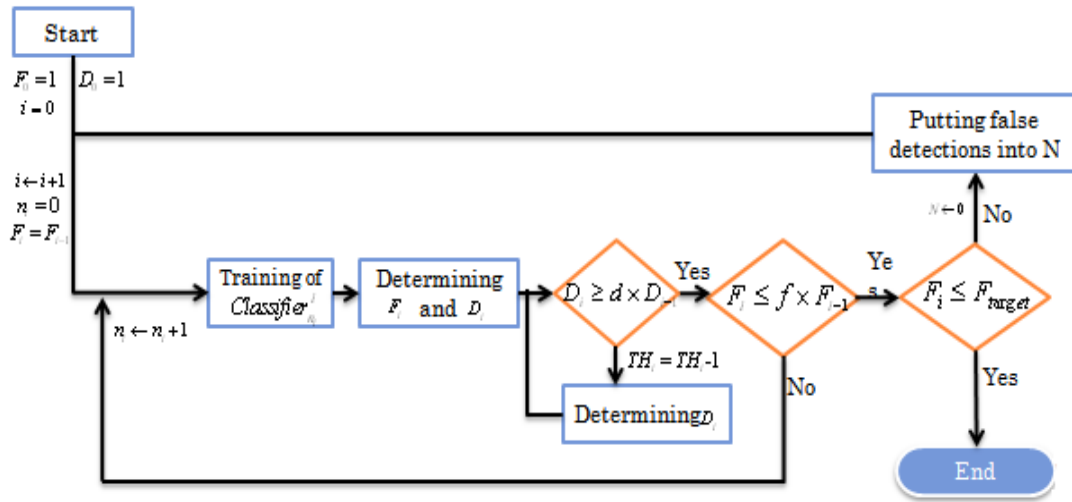
Fig. 2-8: The flow chart of training of cascaded classification

The procedure of our implemented AdaBoost process can be simply described as Fig. 2-9. If m and l are the number of non-faces and faces samples, respectively, and j is the sum of non-faces and faces samples. The initialize weights $w_{i,j}$ for the ith-stage can be defined as $w_{i,j} = \dfrac{1}{2m}, \dfrac{1}{2l}$ for $y_j = 0,1$. The normalized weighted error with respect to the weak classifier can be expressed in the following equation:

$$\varepsilon_i = \min_{f,p,\theta} \sum_j w_{i,j} \left| h(x_j, f, p, \theta) - y_j \right| \qquad (2.3)$$

The updating weights for each iteration are defined in Eq. (2.4), where $e_j = 0$ if

samples x_j is classified correctly, $e_j = 1$ otherwise,

$$w_{i,j} \leftarrow w_{i,j}\beta_i^{1-e_j} \tag{2.4}$$

Also, the final classifier for the ith-stage is defined in the following equation:

$$C(x_j) = \begin{cases} 1, & \alpha_i h(x_j, f, p, \theta) \geq \dfrac{1}{2}\alpha_i \\ 0, & otherwise \end{cases} \tag{2.5}$$

Where $\alpha_i = \log\dfrac{1}{\beta_i}$ and $\beta_i = \dfrac{\varepsilon_i}{1-\varepsilon_i}$.

Start → Normalizing the weights → Selecting the best weak classifier → Updating the weights → End
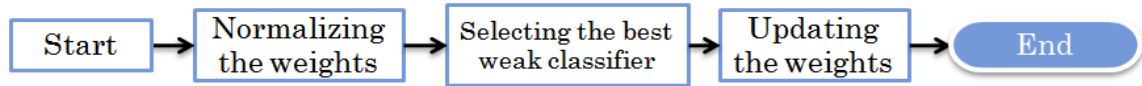
Fig. 2-9: The flow chart of training classification for each stage

2.4 Region-based Clustering

The face detector usually finds more than one face candidates around each face as shown in Fig. 2-10. Here comes the troubled problem when more than two blocks are classified as faces around a single face. So a region-based clustering method is proposed to solve this problem. There are two levels of clustering in this method: local scale clustering and global scale clustering. The local scale clustering is used to cluster the block in the same scale and design a simple filter to determine numbers of blocks within clusters. If numbers of blocks in some cluster are more than one, that cluster will be preserved; otherwise, it will be discarded.
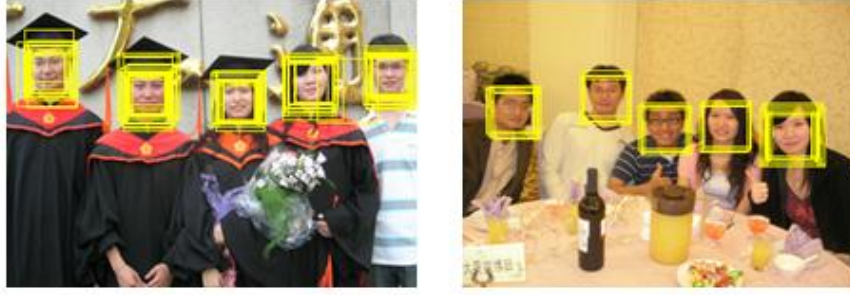
Fig. 2-10: The image after face detecting

$$cluster(x, y) = \begin{cases} 1 & if \ (1) \ and \ (2) \\ 0 & otherwise \end{cases} \tag{2.6}$$

$$(1) \ overlap\_rate(x, y) \geq TH_{overlap\_rate} \tag{2.7}$$

$$(2) \ distance(x, y) \leq TH_{distance} \tag{2.8}$$

Eq. (2.6), Eq. (2.7) and Eq. (2.8) are formulated decision rules of the proposed method. $cluster(x, y) = 1$ means the block x and y are in the same cluster and the regions are completely overlapped. The $overlap\_rate(x, y)$ is the percentage overlapped between two detected region, x and y, and $distance(x, y)$ is the distance of a centers in these two regions. Fig. 2-11 (a) and (b) shows the chart of the overlapped region and the distance of a center of two blocks in the local scale clustering and the global scale clustering respectively. In Fig. 2-11 (a), the two blocks are processed as the same cluster, and in Fig. 2-11 (b), the two blocks are processed as different clusters because the distance of their center does not satisfied with Eq. (2.8).

17

(a)

(b)
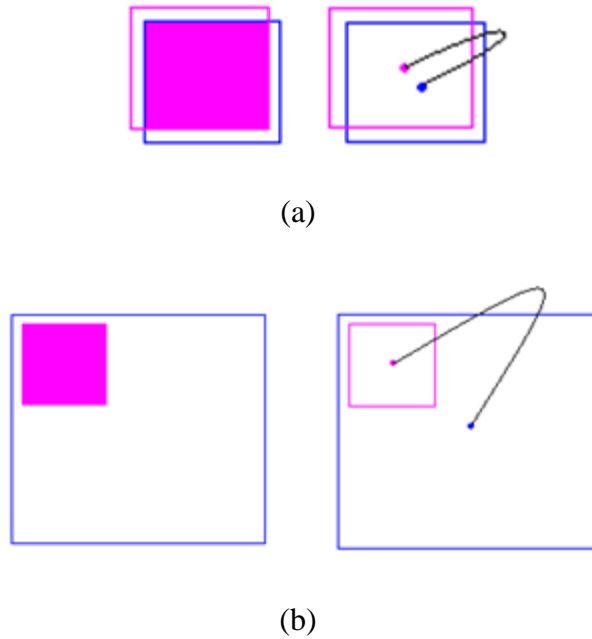
Fig. 2-11: The chart of the overlapped region and the distance of a center of two blocks

in (a) the local scale clustering and (b) the global scale clustering

At the end, the global scale clustering will use the blocks obtained from local scale clustering, and label the face regions by the average size of all available blocks. Some results in the entire region based clustering process for both local-scale and global-scale levels will be shown in Fig. 2-12. From the right image in Fig. 3-10, in fact, only one block will be precisely clustered as a face region after applying our local and global clustering processes even though more than five face candidates are obtained for an image with only five faces.
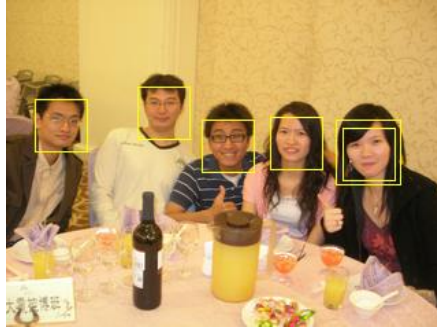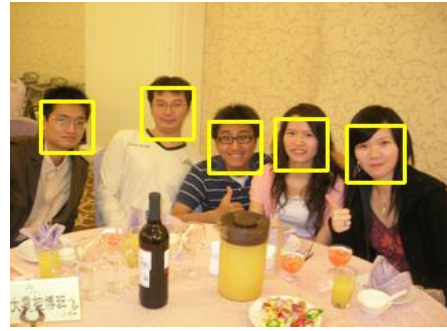
(a)   (b)

Fig. 2-12: (a) The results of the local scale clustering (b) the results of the global scale

clustering
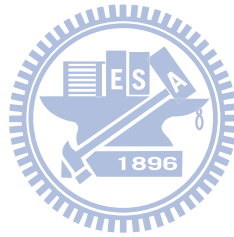
Chapter 3

Age Estimation Algorithm

3.1 Texture Analysis Using Gabor Wavelets

The Gabor wavelets are used for image analysis because their biological relevance and computational properties. Kernels of Gabor wavelets are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit strong characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. The Gabor wavelets transform is *generally acknowledged* to be particularly suitable for image decomposition and representation when the goal is the derivation of local and discriminating features. Moreover, Donato et al [13] have experimentally shown that the Gabor wavelets representation gave better performance for classifying facial actions. In this section, the basics on Gabor wavelets are introduced, the Gabor feature representation of images is described, and a Gabor feature vector for age estimation is derived.

3.2 Gabor Wavelets Transform

A Gabor wavelets $\psi_{\mu,\nu}$ can be defined as follows [14]:

$$\psi_{\mu,\nu}(z) = \frac{\left\|k_{\mu,\nu}\right\|^2}{\sigma^2} e^{-\frac{\left\|k_{\mu,\nu}\right\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}} \right] \tag{3.1}$$

where μ and ν define the orientation and scale of the Gabor kernels, $z = (x, y)$, $\|\cdot\|$ demotes norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{3.2}$$

where $k_\nu = k_{max} / f^\nu$ and $\phi_\nu = \pi\mu/8$. k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. Generally, the Gabor kernels in Eq. (3.1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope a complex plane wave, while the first term in the square brackets in Eq. (3.1), determines the oscillatory part of the kernel and the second term compensates for the DC value. The parameter σ is related to the standard derivation of the Gaussian envelope's width to the wavelength. Fig. 3-1 shows both the real part and imaginary part of Gabor wavelets kernel waveform.

(a)   (b)
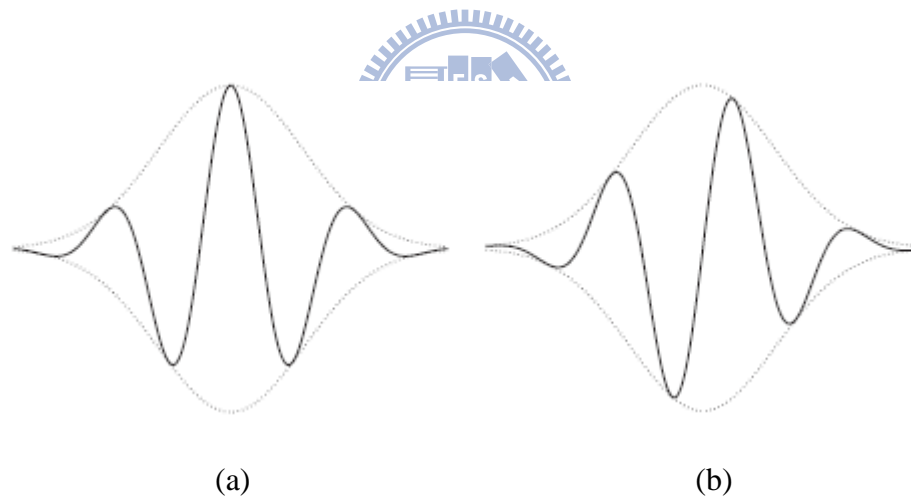
Fig. 3-1: Waveform of Gabor wavelets kernel (a) real part (b) imaginary part
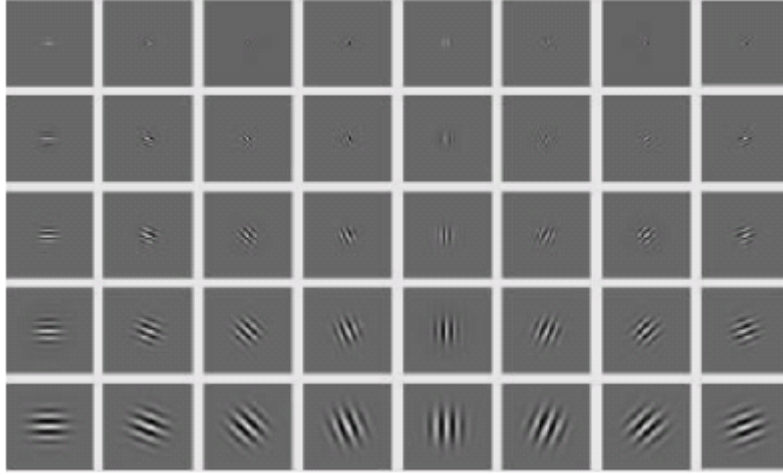
Fig. 3-2: The real part of the 5 x 8 Gabor wavelets

In most cases one would use Gabor wavelets at five different scales, $v \in \{0,\cdots,4\}$, and eight orientations, $\mu \in \{0,\cdots,7\}$ [14][15][16]. Fig. 3-14 shows the real part of the Gabor kernels at five scales and eight orientations and their magnitudes, with the following parameters: $\sigma = 2\pi$, $k_{max} = \pi/2$, and $f = \sqrt{2}$.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by Eq. (3.1). Let $I(z)$ be the gray level distribution of an image, The convolution output of image I and $\psi_{\mu,v}(z)$ is defined as follows

$$O_{\mu,v}(z) = I(z) * \psi_{\mu,v}(z) \tag{3.3}$$

where $z=(x, y)$,and $*$ denotes the convolution operator.

To apply convolution theorem, the Fast Fourier Transform (FFT) is used to derive the convolution output. Eq. (3.4) and Eq. (3-5) are the definition of convolution via FFT.

$$\Im\{O_{\mu,v}(z)\} = \Im\{I(z)\}\Im\{\psi_{\mu,v}(z)\} \tag{3.4}$$

$$O_{\mu,v}(z) = \Im^{-1}\{\Im\{I(z)\}\Im\{\psi_{\mu,v}(z)\}\} \tag{3.5}$$

where $\Im$ and $\Im^{-1}$ denote the Fourier and inverse Fourier transform, respectively.

Fig. 3-3: Sample image and magnitude of 40 convolution outputs

Fig. 3-3 shows the magnitude of convolution outputs of a sample image. The outputs exhibit strong characteristics of spatial locality, scale and orientation selectivity corresponding to those displayed in Fig. 3-2. Such characteristics produce salient local features that are suitable for visual event recognition. From now on, we indicate with $O_{\mu,\nu}(z)$ the magnitude of the convolution outputs.

3.3 Scheme of Gabor Wavelets Features

Generally, convolution results corresponding to all Gabor wavelets are always put together as a whole when Principle Component Analysis (PCA) or other algorithms

which are applied to dimension reduction. In this section, we introduce three different schemes: Parallel Dimension Reduction Scheme (PDRS), Ensemble Dimension Reduction Scheme (EDRS), and Multi-channel Dimension Reduction Scheme (MDRS) [17]. Moreover, we would further compare the performance of them.

Fig. 3-4 shows the principle of PDRS. First, 40 Gabor wavelets features are extracted from each sample. Training each PCA projection matrix in every channels and then combining these features by a voting method that will be described later.
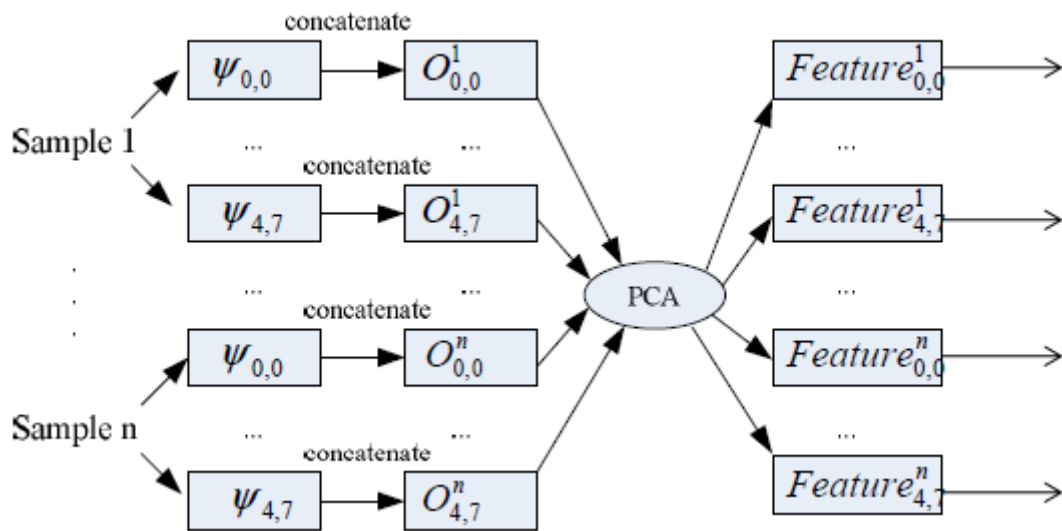
Fig. 3-4: Parallel Dimension Reduction Scheme
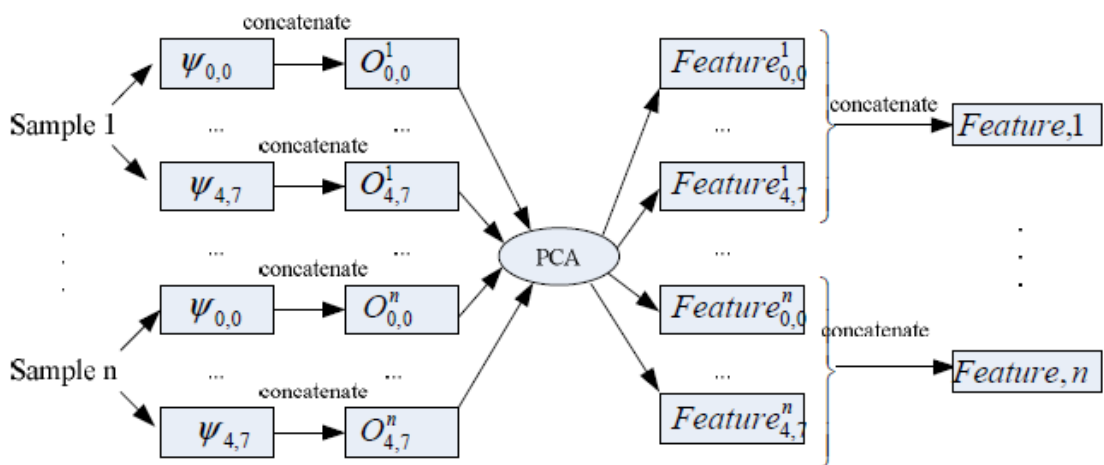
Fig. 3-5: Ensemble Dimension Reduction Scheme
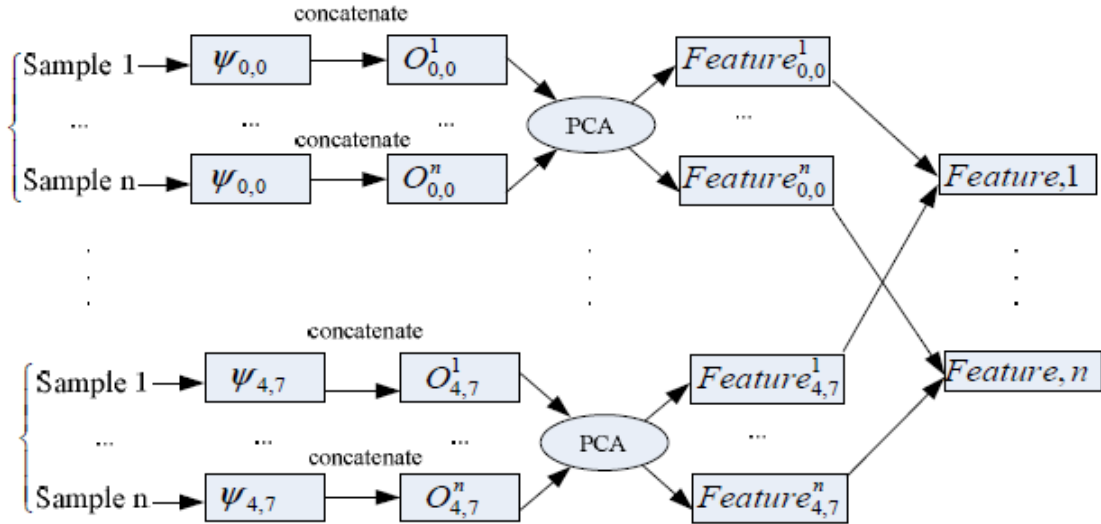
Fig. 3-6: Multi-channel Dimension Reduction Scheme

EDRS is the most common scheme used for Gabor wavelets feature. As shown in Fig. 3-5, the difference between PDRS and EDRS is that the EDRS concatenate 40 Gabor wavelets features instead of using them in parallel. Besides, Xiaodong Li et al [17] proposed the scheme that named as MDRS in 2009. As shown in Fig. 3-6, the main idea of MDRS is training a PCA projection matrix for the same channel between different samples. In [17], Xiaodong Li et al have already proof that MDRS has better performance than EDRS in facial feature extraction using Gabor wavelets transform.

In order to compare the performance of PDRS and MDRS, the K-Nearest Neighbor (KNN) classifier is used for experiment. For PDRS, we use a voting method called Gaussian voting to combine 40 channels. The concept of Gaussian voting is described as using KNN classifier for each channel, and then there are 40 predicted ages. For each predicted age, treat them as the mean value of a Gaussian distribution and count the histogram. The highest peak is the final predicted answer. For MDRS, we use the concatenated feature directly.

The FG-NET Aging Database [18] is adopted for experiments. The database contains 1002 high-resolution color and grey scale face images with large variation of

lighting, pose, and expression. There are 82 subjects (multiple races) in total with the age ranges from 0 to 69 years. We use the mean absolute error (MAE) criterion to evaluate performance of each age estimation. MAE denotes the average of the absolute errors between the estimated ages and ground truth ages. The mathematical function defines as

$$MAE = \sum_{k=1}^{N} \left| \hat{l}_k - l_k \right| / N \tag{3.6}$$

where $\hat{l}_k$ is the ground truth age for the test image k, l_k is the estimated age, and N is the total number of test images. Table 3-1 shows the experimental result of two schemes. The results of experiments demonstrate the MDRS is a more excellent scheme than the PDRS.

Table 3-1: MAE of PDRS and MDRS

Scheme	MAE
PDRS	13.39
MDRS	11.89

3.4 Data Deduction

The dimensionality of the Gabor wavelets feature space is overwhelmingly high, even though the dimension reduction scheme is already apply. In order to reduce the dimension to a low-dimensional space further. We compare four typical dimensionality reduction methods in this section.

(1) Principle Component Analysis (PCA) [19][20]: The concept of PCA is to find a subspace that maximizes the projection variance.

$$p = \arg\max_{\|p\|=1} p^T S p \tag{3.7}$$

where

$$S = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T$$

(3.8)

is the scatter matrix, μ is the mean vector, and N is the number of training sample·

PCA method is mentioned here because it is very popular.

(2) Linear Discriminant Analysis (LDA) [19]: The LDA method is similar to PCA method, the difference is that LDA use class information to improve itself. This method defines two scatter matrixes: between-class scatter matrix

$$S_B = \sum_{i=1}^{C} n_i (\mu_i - \mu)(\mu_i - \mu)^T$$

(3.9)

and within-class scatter matrix

$$S_W = \sum_{i=1}^{C} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

(3.10)

where n_i is the number of class i, μ_i is mean vector of class i, and C is the number of class. The best subspace can defined as

$$p = \arg\max_{p} \frac{|p^T S_B p|}{|p^T S_W p|}$$

(3.11)

(3) Locality Preserving Projections (LPP) [21]: The LPP searches the subspace that preserves essential manifold structure by measuring the local neighborhood distance information. The main idea of LPP is the construction of adjacency undirected graph. Based on the relationship between data define the affinity weight

$$S_{ij} = \exp(-\|x_i - x_j\|^2 / t)$$

(3.12)

Also defines a diagonal matrix

$$D_{ii} = \sum_{j} S_{ij}$$

(3.13)

and a Laplacian matrix

$$L = D - S$$

(3.14)

Then the optimal projection is

$$p = \arg \min_{p} \frac{\left| p^T XLX^T p \right|}{\left| p^T XDX^T p \right|} \tag{3.15}$$

(4) Orthogonal Locality Preserving Projections (OLPP) [22]: The OLPP method produces orthogonal basis functions based on LPP and preserve the metric structure. The optimal projection of OLPP is defined as

$$p = \arg \min_{p^T XDXp=1} \sum_{i=1}^{n} \sum_{j=1}^{n} (p^T x_i - p^T x_j)^2 S_{ij} \tag{3.16}$$

To find out which reduction method of the above is suitable for use in age features from Gabor wavelets, we use the KNN classifier for experiment again. We still use the MAE criterion to evaluate performance. In our experiment, we change the affinity weight of LPP and OLPP to get more detail. Table 3-2 shows the MAE of each reduction method. We can see that OLPP with cosine distance affinity weight has the best performance in age estimation.
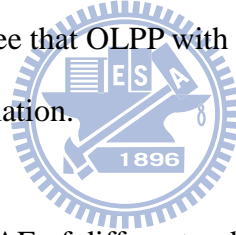
Table 3-2: MAE of different reduction method

Method	Best dimension	MAE
LDA	4096	11.15
LPP	61	10.52
LPP_Heat	254	10.99
LPP_Cosine	45	10.28
OLPP	49	9.24
OLPP_Heat	353	10.99
OLPP_Cosine	43	8.89

3.5 Classification

Support Vector Machines (SVMs) have considerable potential as classifiers of sparse training data which are developed to solve the classification and regression problems. SVMs have similar roots with neural networks, and it demonstrates the well-known ability of being universal approximates of any multivariable function to any desired degree of accuracy. This approach is produced by Vapnik et al. using some statistical learning theory [23][24][25].

3.5.1 Support Vector Machines

Hard-Margin Support Vector Machines

SVM is a way which starts with a linear separable problem. First, we discuss hard-margin SVMs, in which training data are linearly separable in the input space. Then we extend it to the case where training data cannot be linearly separable.

For classification, the objective of SVM is to separate the two classes by a function which is induced from available example. Consider the example in Fig. 3-6, there are two classes of data and many possible linear classifiers that can separate these data. However, only one of them is the best classifier which can maximize the distance between the two classes - margin, this linear classifier is called optimal separating hyperplane.

Given a set of training data $\{\mathbf{x}_i, y_i\}, i = 1, \ldots, m$, where $\mathbf{x}_i \in R^p$, $y_i \in \{+1, -1\}$, where the associate labels are $y_i = 1$ for class1 and -1 for class2. If the data are linearly separable, the decision function can determined by

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - b \tag{3.17}$$
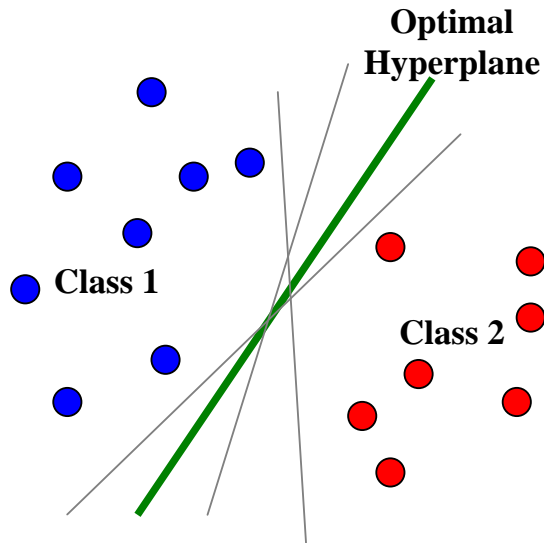
Fig. 3-6: Separating hyperplanes
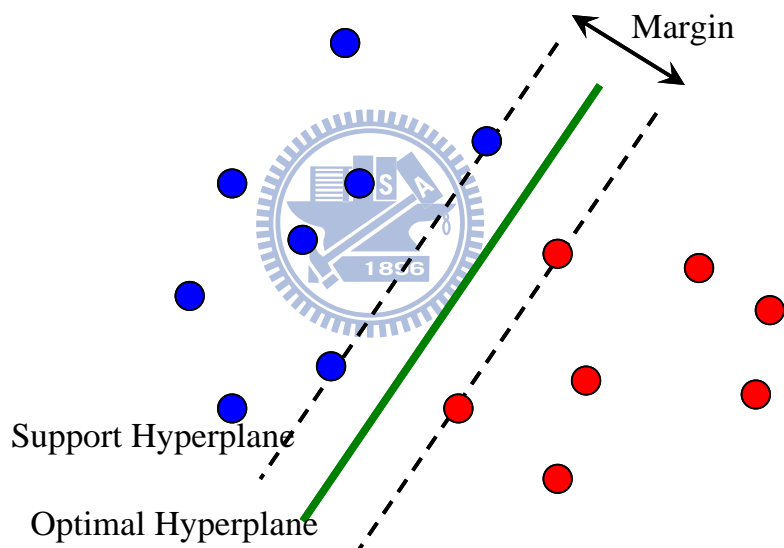

Fig. 3-7: The optimal separating hyperplane

Let

$$\mathbf{w}^T \mathbf{x}_i - b > 0 \quad for \ y_i = +1$$
$$\mathbf{w}^T \mathbf{x}_i - b < 0 \quad for \ y_i = -1$$

(3.18)

The vector $\mathbf{w}$ is a normal vector; it is perpendicular to the hyperplane. The parameter b determines the offset of the hyperplane from the origin along the normal vector $\mathbf{w}$ as shown in Fig. 3-7

Because the training data are linearly separable, without error data satisfying $\mathbf{w}^T\mathbf{x} - b = 0$, we can select two hyperplanes that maximize the distance between two classes. The two hyperplanes include the closest data points which are named support vectors, and also called support hyperplanes. Therefore, the problem can be described by the following equation, after scaling:

$$\begin{aligned} \mathbf{w}^T\mathbf{x} - b \geq +1 \ \ for \ y_i = +1 \\ \mathbf{w}^T\mathbf{x} - b \leq -1 \ \ for \ y_i = -1 \end{aligned} \tag{3.19}$$

The distance between the two support hyperplanes is $2/\|\mathbf{w}\|$. We want to maximize the margin which means to minimize $\|\mathbf{w}\|$. Thus, the problem becomes the following optimization equations:

$$\begin{aligned} choose \ \mathbf{w}, b \ to \ \text{minimize} \ \frac{1}{2}\|\mathbf{w}\|^2 \\ subject \ to \ y_i\left(\mathbf{w}\mathbf{x}_i - b\right) - 1 \geq 0 \ \ \forall i \end{aligned} \tag{3.20}$$

In order to solve the above primal problem of the SVM, we using the method of Lagrange multipliers (Minoux, 1986), and the function will be constructed:

$$L\left(\mathbf{w}, b, \alpha\right) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{m}\lambda_i\left[y_i\left(\mathbf{w}\mathbf{x}_i - b\right) - 1\right] \tag{3.21}$$

λ are the Lagrange multipliers. The Lagrangian has to be minimized with respect to $\mathbf{w}, b$ and be maximized with respect to $\alpha \geq 0$.

$$Larg \ range \ Multiplier \ Condition: \lambda_i \geq 0 \tag{3.22}$$

$$Momplementary \ Slackness: \lambda_i\left[y_i\left(\mathbf{w}^T\mathbf{x}_i - b\right) - 1\right] = 0 \tag{3.23}$$

Minimization with respect to $\mathbf{w}$ and b of Largrangain L is given by:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \implies \sum_{i=1}^{N}\lambda_i y_i = 0 \tag{3.24}$$

$$\frac{\partial L}{\partial b} = 0 \implies \mathbf{w} = \sum_{i=1}^{N}\lambda_i y_i \mathbf{x}_i \tag{3.25}$$

Equation (3.22)-(3.25) are called the KKT conditions (Karush- Kuhn- Tucker conditions). The points of training data which satisfied KKT conditions are the

support vectors, and the solution to the problem is given by,

$$\lambda^* = \arg\min_\lambda \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{k=1}^{m} \lambda_k \tag{3.26}$$

$$\mathbf{w}^* = \sum_{i=1}^{m} \lambda_i y_i \mathbf{x}_i \tag{3.27}$$

$$b^* = \frac{1}{|S|} \sum_{i \in S} y_i - \mathbf{w}^{*T} \mathbf{x}_i \tag{3.28}$$

S is the set of support vectors. Hence, the classifier is simply,

$$f(\mathbf{x}) = \operatorname{sgn}\left(\mathbf{w}^* \mathbf{x} + b^*\right) \tag{3.29}$$

Soft-Margin Support Vector Machines

However, training data are not linearly separable in most situations as shown in Fig. 3-8. There are some training data points on the opposite side. In order to correctly separate the data, a method of introducing an additional cost function associated with misclassification is appropriate as the following equation, where $\xi_i \geq 0$.

$$\begin{aligned} \mathbf{w}^T \mathbf{x} - b &\geq +1 - \xi_i \quad \textit{for } y_i = +1 \\ \mathbf{w}^T \mathbf{x} - b &\leq -1 + \xi_j \quad \textit{for } y_i = -1 \end{aligned} \tag{3.30}$$

In this situation, the classifier becomes more powerful when the residual value ξ_i becomes smaller, thus we need to minimize the cost function.

$$Cost = C \left(\sum_i \xi_i \right)^k \tag{3.31}$$

Therefore, the problem becomes:

$$\begin{aligned} &\textit{choose } \mathbf{w}, b \textit{ to minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ &\textit{subject to } y_i \left(\mathbf{w} \mathbf{x}_i - b \right) - 1 \geq 0 \quad \forall i \\ &\qquad\qquad \xi_i \geq 0 \quad \forall i \end{aligned} \tag{3.32}$$
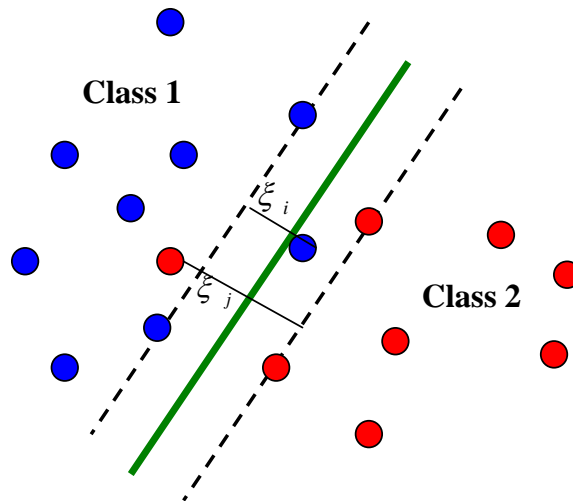
Fig. 3-8: Inseparable case in a two-dimensional space

$$x \rightarrow \phi(x)$$
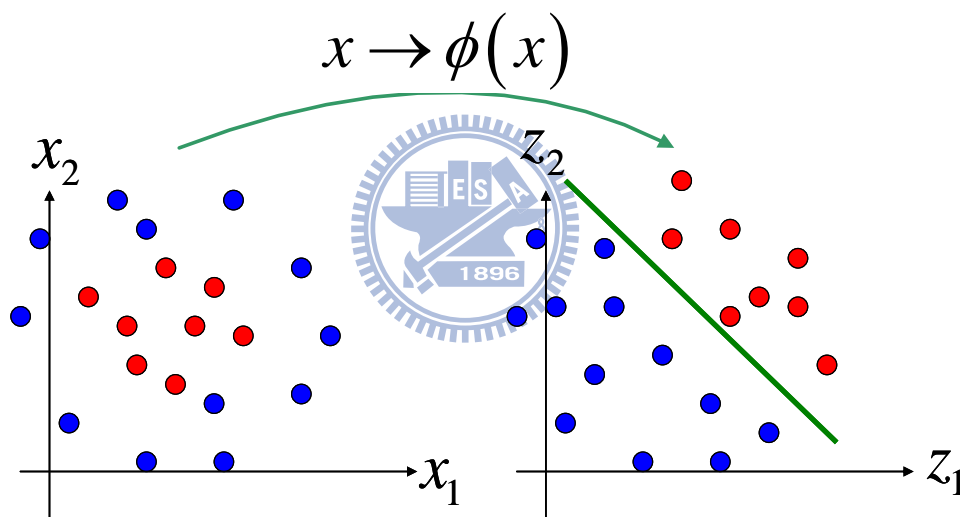
Fig. 3-9: Transformation of feature space

Mapping to a High-Dimensional Space

If the training data are not linearly separable, we can enhance the linear separability in a feature space by mapping the data from the input space into the high-dimensional feature space. Here we show an example in Fig. 3-9. The resulting algorithm is formally similar except that every dot product is replaced by a non-linear kernel function k. It allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space.

In the following are some kernels that are in common use with support vector machines.

Linear kernels: $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$

Polynomial kernels: $k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}')^d$

Radial basis function kernels: $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma \left\| \mathbf{x} - \mathbf{x}' \right\|^2\right),\ for\ \gamma > 0$

Sigmoid: $k(\mathbf{x}, \mathbf{x}') = \tanh\left(k\mathbf{x}^T \mathbf{x}' + c\right)$
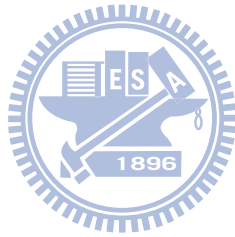
Chapter 4

Experimental Results

In this chapter, the results of age estimation will be presented. Our algorithm was implemented on the platform of PC with Intel Pentium Duo 2GHz and 4GB RAM. Matlab and Borland C++ Builder are our complier and operated on Windows XP. All of our testing inputs are images from file.

In section 4.1, we will introduce the dataset used in the experiments. Besides, performance analysis and comparison between difference algorithms are demonstrated in section 4.2. The criterions of performance were evaluated by mean absolute error and cumulative score.
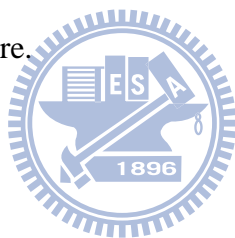
4.1 Dataset

The database that we adopt for age estimation experiments is the FG-NET Aging Database [18]. This database is a public available age database and it contains 1002 high-resolution color or grey scale face images with large variation of lighting, pose, and expression. There are 82 subjects (multiple races) in total with the age ranges from 0 to 69 years old. Fig. 4-1 shows a serial sample images from the same subject with different age in the database

Fig. 4-1: Sample images in different ages from the same subject

4.2 Results

To evaluate the age estimation performance, the face area in each image was located by face detector described in chapter 2 first. A leave-one-person-out (LOPO) test scheme is used on the experiments. Each face image is cropped and resized to 64 by 64 pixels in size and transformed color information to 256 gray level. We use the classifier of SVMs with parameters of RBF kernel where cost c is 0.5 and gamma g is 0.0078125 for the FG-NET database. Our main focus is the new feature based on Gabor wavelets.

The performance of age estimation can be measured by two different measures: the mean absolute error (MAE) and the cumulative score (CS). The MAE is defined as the average of the absolute errors between the estimated ages and the ground truth ages,

$$MAE = \sum_{k=1}^{N} \left| \hat{l}_k - l_k \right| / N \tag{4.1}$$

where $\hat{l}_k$ is the ground truth age for the test image k, l_k is the estimated age, and N is the total number of test images. The MAE measure was used previously in [1]-[9]. The cumulative score is defined as

$$CS(j) = N_{e \leq j} / N \times 100\% \tag{4.2}$$

where $N_{e \leq j}$ is the number of test images on which the age estimation makes an

absolute error no higher than j.

Table 4-1 shows the experimental results. We compare our results with all previous methods reports on the FG-NET age database. It turns out our Gabor-OLSS method has the MAEs of 8.43 and 5.71 years for using KNN and SVM, which are explicitly smaller than most of previous results under the same experimental protocol. Our method brings about 16% deductions of MAEs over the result of AGES [1]. In Table 4-1, we can find the LARR [3] method has a better MAE of 5.07 than ours. As we mentioned before, our purpose is to build a "full-automatic" age estimation system. The LARR method use the AAM features FG-NET provided directly, that is it need human involved for aligning the feature points or aligning them automatically. In our survey, there still isn't an efficient method can align feature points fast and correctly. For applications, the LARR method may take large effort on aligning features points. We implement the BIF method. The result of BIF implemented is quite bad; it has the MAE of 10.32. Furthermore, the BIF method requires a large amount of time when extracting the aging features. Compare to our system, the BIF method takes more than twice of our extraction time. Our method brings the performance of feature extraction about 12 to 15 images per second.

Table 4-1: MAE of different methods

Method	MAE
WAS [1]	8.06
AGES [1]	6.77
MLPs [26]	6.98
LARR [3]	5.07
BIF implemented	10.32
Ours (KNN)	8.43

Ours (SVM)	5.71

The comparisons of cumulative scores are shown in Fig. 4-2. Our Gabor-OLPP method performs much better than WAS and MLPs methods. The method of AGES is close to our Gabor-OLPP method in low age error levels, but worse than Gabor-OLPP in high levels.
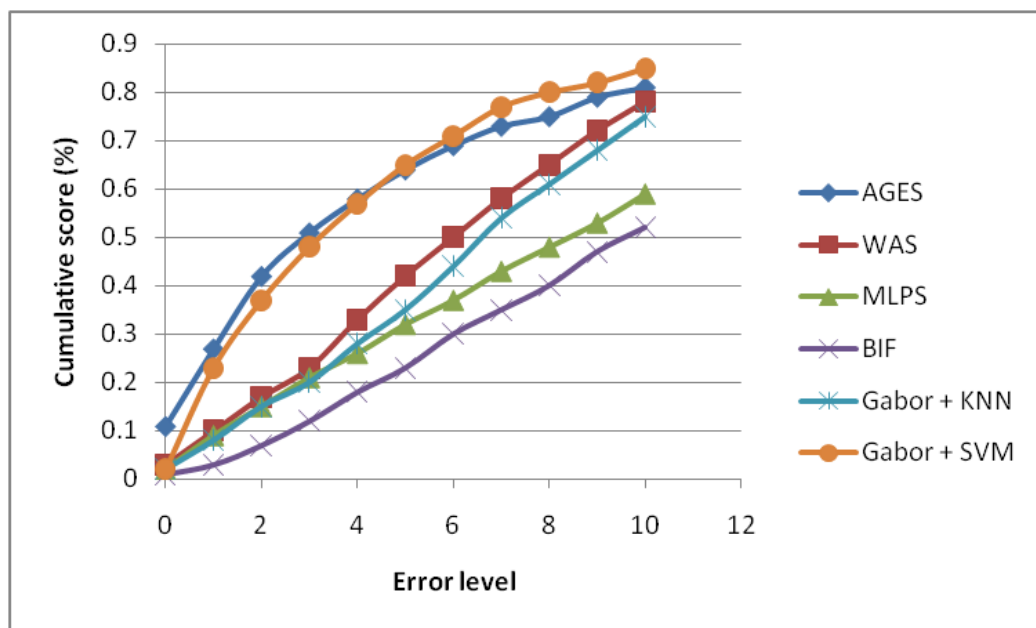

Fig. 4-2: CS of each method

Chapter 5

Conclusions and Future Work

In this thesis, we proposed a new framework for automatic age estimation of face image. Gabor wavelets transform is first introduced to age estimation for achieving real-time and full automatic aging feature extraction; SVMs have considerable potential as classifiers of sparse training data and provide robust generalization ability.

Most of references are only using PCA to reduce the dimensionality of Gabor wavelets features. But it has not enough efficiency when we use general Gabor wavelets features directly. By trading off efficiency and accuracy of classification, former authors usually try to select the features they need only, rather than just use all features. Therefore, data reduction methods are more convenience for "selection" relatively for select the features we want. We compare four different typical data reduction methods; OLPP gives the lowest dimensionality of feature vectors and the best discrimination in further extraction.

Although a better age estimation performance such as computational speed and MAE are obtained by Gabor-OLPP-SVM method than most of methods, there are some details we can improve in the future. Gabor wavelets feature selection is the main issue. OLPP already gives a good performance in reduction, but we can select the most important features from forty Gabor wavelets features before combining them. This may improve the performance of whole system. We can also add additional local features. Since the Gabor wavelets affect on full image, we add other local features such as edges or contours may gives a better accuracy rate on age estimation. The image size we handled is 64 by 64 pixels, this size is a bottleneck

when using face detection by Viola-Jones classifier. So finding the balance between image size and the lowest needed resolution for Gabor wavelets features is another important issue for improving our system.
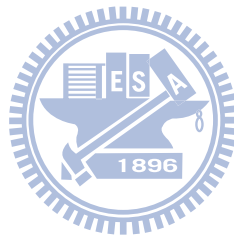
Reference

[1] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," *In ACM Conf. on Multimedia*, pages 307–316, 2006.

[2] Xin Geng, Zhi-Hua Zhou, and Smith-Miles, K., "Automatic age estimation based on facial aging patterns," *IEEE Trans. on PAMI*, 29(12): 2234–2240, 2007.

[3] Guodong Guo, Yun Fu, Dyer, C.R., and Huang, T.S., "Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression," *IEEE Trans. on Image Processing*, 17(7): 1178-1188, 2008.

[4] Guodong Guo, Yun Fu, Huang, T.S., and Dyer, C.R., "Locally Adjusted Robust Regression for Human Age Estimation," *IEEE Workshop on Applications of Computer Vision*, pages 1-6, 2008.

[5] Y. Kwon and N. Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, 74(1): 1–21, 1999.

[6] Jun-Da Txia, Chung-Lin Huang, "Age Estimation Using AAM and Local Facial Features," *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 885-888, 2009.

[7] Shuicheng Yan, Xi Zhou, Ming Liu, Hasegawa-Johnson, M., and Huang, T.S., "Regression from patch-kernel," *IEEE Conference on CVPR*, pages 1-8, 2008.

[8] Guodong Guo, Guowang Mu, Yun Fu, Huang, T.S., "Human age estimation using bio-inspired features," *IEEE Conference on CVPR*, pages 112-119, 2009.

[9] Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T., "Robust Object Recognition with Cortex-Like Mechanisms," *IEEE Trans. on PAMI*, 29(3):

411–426, 2007.

[10] I-Shan Chen, Chin-Teng Lin, "Multi-client Identification System using Adaptive Probabilistic Model," CS, NCTU.

[11] Viola Paul and J. Jones. Michael, "Robust Real-Time Face Detection," *International Journal of Computer Vision 57(2)*, 137-154, 2004.

[12] C.P. Papageorgiou, M. Oren, and T. Poggio, "A General Framework for Object Detection," *International Conference on Computer Vision*, pp. 555-562, Jan. 1998.

[13] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 974–989, Oct. 1999.

[14] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779, 1997.

[15] C. Liu and H. Wechsler, "Gabor feature based classification using enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, pp. 467–476, 2002.

[16] C. Liu, "Gabor-based kernel pca with fractional power polynomial models for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 572–581, 2004.

[17] Xiaodong Li, Shumin Fei, and Tao Zhang, "Novel Dimension Reduction Method of Gabor Feature and Its Application to Face Recognition, " International Congress on Image and Signal Processing, 2009. CISP '09. 2nd, Page(s): 1 − 5, 2009

[18] The FG-NET Aging Database [Online]. Available: http://www.fgnet.rsunit.com/

[19] Belhumeur, P. N., J. P. Hespanha, et al. (1997). "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7): 711-720.

[20] R. O. Duda, E. H. Peter, and G. S. David, *Pattern Classification*, 2nd ed. New York: Wiley Interscience, 2000.

[21] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Niyogi P., Hong-Jiang Zhang (2005). "Face recognition using Laplacianfaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(3): 328-340.

[22] Deng Cai, Xiaofei He, Jiawei Han, Hong-Jiang Zhang (2006). "Orthogonal Laplacianfaces for Face Recognition." *IEEE Transactions on Image Processing* 15(11): 3608-3614.

[23] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," in *Proc. IGARSS*, Toulouse, France, July 21–25, 2003.

[24] Abe, Shigeo, "Support Vector Machines for Pattern Classification," London: Springer-Verlag London Limited, 2005.

[25] L Wang ed., "Support Vector Machines: Theory and Applications," New York: Springer, Berlin Heidelberg, 2005.

[26] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 621–628, Feb. 2004.